VIII kadencja

KANCELARIA SEJMU
Biuro Komisji Sejmowych

PEŁNY ZAPIS PRZEBIEGU POSIEDZENIA

■ **Komisji Cyfryzacji, Innowacyjności i Nowoczesnych Technologii (nr 115)**

z dnia 22 listopada 2018 r.

Pełny zapis przebiegu posiedzenia

Komisji Cyfryzacji, Innowacyjności i Nowoczesnych Technologii (nr 115)

22 listopada 2018 r.

Komisja Cyfryzacji, Innowacyjności i Nowoczesnych Technologii, obradująca pod przewodnictwem posła **Pawła Pudłowskiego (N)**, przewodniczącego Komisji, rozpatrzyła:

– informację Ministra Cyfryzacji o otwieraniu danych publicznych w ramach projektu POPC „Otwarte dane – dostęp, standard, edukacja”,

– informację Ministra Cyfryzacji o projekcie dyrektywy o ponownym wykorzystywaniu informacji sektora publicznego.

W posiedzeniu udział wzięli: **Karol Okoński** sekretarz stanu w Ministerstwie Cyfryzacji wraz ze współpracownikami, **Joanna Karczewska** przewodnicząca Komisji Rewizyjnej Stowarzyszenia ISACA, **Włodzimierz Marciński** prezes Polskiego Towarzystwa Informatycznego, **Melanie Raczek** członek Konfederacji Lewiatan, **Witold Staniszkis** przedstawiciel Polskiej Izby Informatyki i Telekomunikacji oraz **Jerzy Żurek** przedstawiciel Rady Głównej Instytutów Badawczych.

W posiedzeniu udział wzięli pracownicy Kancelarii Sejmu: **Ewa Gast** oraz **Anna Ornat** – z sekretariatu Komisji w Biurze Komisji Sejmowych.

Przewodniczący poseł Paweł Pudłowski (N):

Dzień dobry. Witam państwa bardzo serdecznie. Witam państwa posłów. Witam pana ministra i państwa z ministerstwa. Witam wszystkich gości. Otwieram 115 posiedzenie Komisji Cyfryzacji, Innowacyjności i Nowoczesnych Technologii. Porządek dzisiejszego posiedzenia przewiduje rozpatrzenie informacji Ministra Cyfryzacji o otwieraniu danych publicznych w ramach projektu POPC „Otwarte dane – dostęp, standard, edukacja” oraz – drugi punkt – informacji Ministra Cyfryzacji o projekcie dyrektywy o ponownym wykorzystywaniu informacji sektora publicznego. Powyższy porządek państwo posłowie otrzymali.

Czy są uwagi do porządku dziennego? Nie słyszę uwag. Stwierdzam, że Komisja przyjęła porządek dzienny posiedzenia bez zmian.

Realizacja porządku dziennego. Proponuję, aby oba punkty zostały zreferowane jednocześnie, a potem będzie wspólna dyskusja odnośnie do obu punktów. Jeśli nie ma sprzeciwu, to proszę uprzejmie pana sekretarza stanu w Ministerstwie Cyfryzacji Karola Okońskiego o przedstawienie informacji.

Sekretarz stanu w Ministerstwie Cyfryzacji Karol Okoński:

Panie przewodniczący, Wysoka Komisjo, na dzisiejszym posiedzeniu chcielibyśmy przedstawić zarówno główne założenie, jak i stan prac nad dyrektywą dotyczącą ponownego wykorzystania informacji sektora publicznego, tzw. dyrektywą PCI – *public sector information* – a oprócz tego, informację o dwóch projektach, które realizujemy w ramach otwierania danych publicznych w Polsce, finansowanych ze środków unijnych. Do przedstawienia tego materiału poproszę zaraz panią dyrektor Agatę Miazgę, która opiekuje się oboma tematami w Ministerstwie Cyfryzacji. Jednak, zanim oddam głos, jedna, myślę że ważna, informacja odnośnie do samej dyrektywy, że w tym momencie zdaje się, iż na poziomie państw członkowskich i Rady Unii Europejskiej udało się osiągnąć kompromis.

Został 7 listopada wypracowany wspólny mandat rady prezydencji do dalszych prac nad tą dyrektywą. Zakładamy, że w I kwartale, czyli już po Nowym Roku, zakończy prace Parlament Europejski, rozpoczną się trilogi i ambitny cel jest taki, żeby ta dyrektywa została uchwalona do końca I kwartału przyszłego roku. Jednak z racji tego, że jest to dyrektywa, a nie rozporządzenie, będzie potem dwuletni okres na implementację do polskiego porządku prawnego. Tak więc, z perspektywy samej dyrektywy, na efekty będziemy musieli trochę poczekać, ale oczywiście pamiętajmy, że część z tych postulatów już de facto realizujemy, czyli, są ujęte w polskim prawie, czy są realizowane w naszych projektach. A teraz, jeśli pan przewodniczący pozwoli, oddałbym głos pani dyrektor.

Dyrektor Departamentu Otwartych Danych i Rozwoju Kompetencji w MC Agata Miazga:
Panie przewodniczący, szanowni państwo, na początek opowiem o tej dyrektywie, a potem o samym projekcie „Otwarte dane". Przybliżę państwu cele nowej dyrektywy.

Głównym celem jest wzmocnienie gospodarki opartej na danych, także zwiększenie skali powstawania innowacyjnych produktów i usług, które opierają się na tych danych oraz zapewnienie dostępu w czasie rzeczywistym do danych dynamicznych – to jest ważna kwestia, która jest podkreślona w dyrektywie – za pośrednictwem interfejsu programistycznego API. Wreszcie chyba podstawowy cel dyrektywy, czyli zwiększenie podaży tych danych publicznych, które mają wysoką wartość, duży potencjał ponownego wykorzystywania, w tym – nowość – danych pochodzących z sektora przedsiębiorstw publicznych, organizacji prowadzących badania naukowe, instytucji finansujących badania ze środków publicznych.

Dyrektywa rozszerza zakres ponownego wykorzystywania informacji sektora publicznego o dwie kategorie danych. Po pierwsze, są to dane będące w posiadaniu przedsiębiorstw publicznych – tych, które prowadzą działalność w sektorach gospodarki wodnej, energetyki, transportu i usług pocztowych.

Drugą kategorią danych, którą dyrektywa rozszerza zakres ponownego wykorzystywania, są dane badawcze. Chciałabym opowiedzieć w skrócie o tych rodzajach danych.

Jeżeli chodzi o dane przedsiębiorstw publicznych, to tutaj pojawiła się taka definicja przedsiębiorstwa publicznego, że jest nim przedsiębiorstwo, na które organy sektora publicznego mogą wywierać dominujący wpływ. Ten dominujący wpływ może być z racji tego, że państwo jest jego właścicielem, posiada odpowiednie udziały finansowe albo po prostu ma ten dominujący wpływ z mocy przepisów. Tak więc, jak już mówiłam, przedsiębiorstwa publiczne objęte tą dyrektywą, to te, które działają w sektorach gospodarki wodnej, energetyki, transportu i usług pocztowych.

Natomiast, jeżeli chodzi o dane badawcze, są to dane, nie publikacje naukowe – to jest ważne rozróżnienie – które są gromadzone w trakcie działalności badawczo-naukowej. To są te dane badawcze, które już zostały opublikowane w jakimkolwiek repozytorium cyfrowym. Te dane, które już są dostępne, ale teraz będą miały jeszcze ten dodatkowy wymiar, czyli będą udostępnione do ponownego wykorzystywania. Oprócz tego, państwa członkowskie będą zobowiązane do przyjęcia polityki krajowej na rzecz powszechnej dostępności danych badawczych pochodzących z badań naukowych, tych, które są finansowane ze środków publicznych, czyli tzw. *open access*.

Dyrektywa wzmacnia znaczenie zautomatyzowanego dostępu do danych. Chodzi tu o dostęp do danych, po pierwsze, dynamicznych, czyli tych, które często się zmieniają. To są np. dane meteorologiczne, o zanieczyszczeniu powietrza albo rozkłady jazdy. Takie dane dynamiczne powinny być udostępniane natychmiast po ich zgromadzeniu, za pośrednictwem API. Dyrektywa jednak trochę łagodniej podeszła do tematu. To znaczy, że jeśli udostępnienie tych danych natychmiast po zgromadzeniu wiązałoby się z dużymi kosztami lub pojawiłyby się trudności techniczne, to takie dane dynamiczne mogą być udostępniane w późniejszym możliwym terminie, czyli nie natychmiast po zgromadzeniu.

Druga kategoria danych, które mają być udostępniane w sposób automatyczny, to dane o wysokiej wartości. To jest bardzo ważna kategoria wprowadzona w tej dyrektywie. Są to te informacje sektora publicznego, które mają duży potencjał do ponownego wykorzystywania w produktach i usługach bazujących właśnie na danych.

Zbiór danych o wysokiej wartości zostanie określony przez komisje w drodze aktu wykonawczego. Pojawi się katalog danych, które będą musiały być udostępniane nieodpłatnie w formacie nadającym się do odczytu maszynowego i za pośrednictwem API. Zatem, bardzo ważny będzie ten katalog danych, które zostaną określone w drodze aktu wykonawczego.

Z tego udostępniania, jako dane o wysokiej wartości, wyłączone zostaną dane będące w posiadaniu bibliotek, archiwów i muzeów, a także przedsiębiorstw publicznych, jeśli udostępnienie takich danych doprowadziłoby do znacznego zakłócenia konkurencji na rynkach. Wyłączone na 2 lata będą także te zbiory danych organów sektora publicznego, które uzyskują dochody w celu pokrycia znacznej części kosztów związanych w wykonywaniem zadań publicznych. Czyli, jeśli instytucja czerpie dochody z np. udostępniania danych, to będzie miała możliwość odłożenia tego obowiązku na 2 lata.

Tutaj przykładowo pokazaliśmy państwu, jak mogą wyglądać te dane o wysokiej wartości. Oczywiście, to jeszcze nie jest propozycja komisji. To jest komunikat komisji z 2014 r., który prawdopodobnie będzie punktem wejścia do prac komisji nad tym zakresem danych wysokich wartości. Jak widzimy, są to dane w dużej mierze już udostępniane u nas w kraju, ale niekoniecznie przez API i niekoniecznie bezpłatnie. Mam tu na myśli głównie dane geoprzestrzenne czy rejestry spółek, które do tej pory nie wszystkie są udostępniane w Polsce, w sposób otwarty. Jest to więc duże wyzwanie.

Konsekwencje dla Polski. Konsekwencje dyrektywy dla Polski są przede wszystkim szansami. Jest to duża szansa dla polskich przedsiębiorstw, które opierają swoją działalność na danych publicznych, wykorzystują je w aplikacjach czy usługach. Będzie to też się wiązało z pewnymi dodatkowymi obowiązkami, takimi, że polska administracja będzie musiała się dostosować do tych wymogów, chociażby poprzez budowę API do niektórych systemów, a także nieco inaczej podchodzić do udostępniania danych publicznych. Wreszcie, konieczne będą zmiany w prawie, czyli ustawa o ponownym wykorzystywaniu informacji sektora publicznego będzie musiała być dostosowana do nowej dyrektywy. Właściwie to w dużym skrócie cała informacja o dyrektywie.

Może przejdę teraz do projektu, który, tak naprawdę, częściowo odpowiada na te wyzwania jakie w dyrektywie się znalazły. Projekt „Otwarte dane – dostęp, standard, edukacja", rozpoczęliśmy w czerwcu 2017 r. Jest to projekt indykatywny. To oznacza, że można powiedzieć, iż został uznany za systemowy, wprowadzający ważne zmiany dla właściwie całej administracji centralnej. Jego głównym celem jest poprawa dostępności i jakości danych publicznych, a także zwiększenie możliwości ich ponownego wykorzystywania.

Ten cel jest niemalże tożsamy z celem działania 2.3.1 Programu Operacyjnego Polska Cyfrowa. Dla nas projekt ten jest zarazem wsparciem dla realizacji zadań z Programu Otwierania Danych Publicznych, czyli naszego głównego dokumentu strategicznego, który ma na celu otwieranie danych publicznych administracji centralnej, a który nie przewiduje środków finansowych na realizację tych zadań.

Projekt zakończy się w maju 2020 r. Uczestniczą w nim, oprócz Ministerstwa Cyfryzacji, trzej partnerzy – Ministerstwo Finansów, Główny Urząd Statystyczny oraz Narodowy Fundusz Zdrowia. Kosztuje on 23,5 mln zł. Tak jak już wspomniałam, dofinansowany jest z POPC, z działania 2.3.1. Już teraz przewidujemy jego kontynuację, ale o tym zaraz. Teraz opowiem trochę więcej o nowym projekcie.

Projekt ma 3 główne obszary działania – dostęp, standard i edukacja. W ramach dostępu, otwieramy przez API 6 baz danych. Czyli, można powiedzieć, że może w sposób nie do końca dynamiczny, ale najbardziej otwarty, udostępniamy dane z sześciu baz danych, które są kluczowe dla obywateli i przedsiębiorców. Są to bazy danych NFZ – 3 bazy, baza GUS, MC oraz MF. Drugim działaniem w ramach tego zwiększania dostępu do danych jest rozbudowa i dostosowanie do potrzeb odbiorców portalu dane.gov.pl.

Drugim kierunkiem działań są standardy. Opracowaliśmy już – to jest działanie zakończone – standardy otwartości danych w trzech wymiarach: regulacji prawnych, technicznym oraz bezpieczeństwa.

Wreszcie edukacja, bardzo ważne działanie w projekcie. To cykl szkoleń dla administracji, opracowanie materiałów edukacyjnych oraz działania informacyjno-promocyjne,

których celem jest poprawa jakości i otwartości tych danych, które są udostępniane przez administrację centralną, a także zwiększenie świadomości urzędników nt. korzyści z udostępniania danych do ponownego wykorzystywania.

Przez ostatnie 18 miesięcy udało się nam już część z tych zadań zrealizować. Przede wszystkim – jak już wspomniałam – opracowaliśmy standardy otwartości danych. Standardy są 4. Pierwszy, to standard prawny, w którym m.in. identyfikujemy bariery w udostępnianiu danych publicznych w sposób otwarty, takie, które pojawiają się w aktach prawnych. Opisujemy dokładnie filary otwartości danych, a także wzorzec otwartości danych – prawny wzorzec otwartości. Wskazujemy też zalecenia dotyczące ram prawnych, które pozwolą na powszechny i swobodny odstęp do danych i ich ponowne wykorzystywanie.

Drugi, to standard bezpieczeństwa, który określa minimalne czynności dotyczące udostępniania danych do ponownego wykorzystywania w sposób bezpieczny. Określamy tam także grupy danych, które podlegają nieograniczonemu ponownemu wykorzystywaniu, a także te, które wymagają anonimizacji lub pseudonimizacji. Mówimy też o tym, jak dobrać określone techniki anonimizacji do grup danych.

Standard techniczny, to wymagania techniczne dla danych publicznych. Tutaj skupiliśmy się na pięciu poziomach otwartości danych powszechnie stosowanych wg schematu Tima Berners-Lee. Opisaliśmy też dokładnie, jak powinny wyglądać formaty danych, jakie wymagania powinny spełniać i jakie są przykłady prawidłowego przygotowania danych. Określiliśmy też, jak należy przygotowywać pewne typy danych, takie jak dane adresowe czy data.

Wreszcie, standard API, który – jak można powiedzieć – jest rozwinięciem standardu technicznego. Tam przedstawiliśmy minimalne rekomendowane zalecenia dotyczące API, które udostępniają dane z baz administracji publicznej. Tam jest m.in. mowa o architekturze usług sieciowych REST, jest opis składni, obsługa zdarzeń http, wiele szczegółów technicznych, o których nie będę tutaj mówić. Wszystkie standardy znajdą państwo na naszej stronie dane.gov.pl. Są dostępne w formatach otwartych.

Drugie działanie w projekcie to przebudowa portalu dane.gov.pl. Ona jest w trakcie. Został nam jeszcze rok prac nad portalem. Na razie udało się nam zmienić, poprawić, usprawnić wyszukiwarkę na stronie, która jest właściwie centralnym elementem portalu i umożliwia swobodne wykorzystanie danych. Poprawiliśmy także, a właściwie przygotowaliśmy od nowa, intuicyjny interfejs, który jest zgodny z WCAG 2.0. Opracowaliśmy także nowe API, ustandaryzowane, zgodne z naszym standardem. Nasz portal zyskał też nową domenę dane.gov.pl. Poprzednio było to danepubliczne.gov.pl. W portalu udostępniamy dane bezpłatnie. Są one na bieżąco aktualizowane. Udostępnia je już ponad 120 instytucji.

Działania szkoleniowe to kolejny wymiar naszego projektu. Rozpoczęły się one pod koniec zeszłego roku, a będą trwały do końca przyszłego. Są to szkolenia dla urzędników zatrudnionych w administracji centralnej.

Przewidzieliśmy dwa moduły. Moduł podstawowy, w którym mówimy o aspektach prawnych, czyli o ustawie, o dostępie i ustawie o ponownym wykorzystywaniu, a także o podstawowych wymaganiach i zaleceniach technicznych dotyczących przygotowywania danych do ponownego udostępniania.

Jest też moduł zaawansowany, który rozpoczyna się w tym roku. Tam będziemy szkolić ze standardów. Ze wszystkich tych czterech standardów, o których wspomniałam.

Dotąd udało się nam przeszkolić 119 pracowników administracji, ale te wskaźnik cały czas rośnie. Docelowo będzie to ok. 300 osób.

Dodatkowo prowadzimy szkolenia sprofilowane dla dysponentów danych, które przygotowujemy na potrzeby konkretnych instytucji. Dotychczas udało się nam przeszkolić urzędników z pięciu instytucji. Te szkolenia organizowane są przez członków zespołu projektowego z Ministerstwa Cyfryzacji. Są one wprost skrojone pod potrzeby konkretnych urzędników. Prowadzimy także szkolenia dla redaktorów, którzy udostępniają dane na portalu dane.gov.pl.

Kolejne, bardzo ważne działanie w projekcie to dobudowa API do sześciu baz danych. Pierwsza baza, którą udało się nam już właściwie udostępnić przez API – API wkrótce

zostanie opublikowane, ale już jest gotowe – to bank danych lokalnych GUS, czyli największa baza danych o regionach i właściwie każdej dziedzinie życia.

Drugim partnerem, który udostępnia dane przez API w ramach naszego projektu, jest NFZ. Tutaj mamy bazę „kolejki", czyli czas oczekiwania na świadczenia medyczne. Tutaj prace nad API właściwie już się zakończyły. Trwają jeszcze prace nad API do bazy „umowy" – to są umowy zawierane przez NFZ z placówkami medycznymi – a także do statystyk Jednorodnych Grup Pacjentów, czyli gigantycznej bazy wiedzy o hospitalizacjach, zachorowaniach i wszelkich kwestiach dotyczących statystyk w dziedzinie zdrowia.

Trzecim dysponentem danych, który udostępnia dane przez API, jest Ministerstwo Cyfryzacji. Trwają w tej chwili prace nad budową API do Centralnej Ewidencji Pojazdów. Docelowo będziemy także udostępniać, oprócz danych technicznych pojazdów, statystyczne dane o kierowcach.

Ostatnią instytucją jest Ministerstwo Finansów. Tu pracujemy nad API do BESTI, czyli systemu udostępniającego dane o dochodach i wydatkach samorządów.

Trwają także inne działania. Mamy w projekcie zaplanowane działania informacyjno-promocyjne. W zeszłym miesiącu odbył się „Hackathon – nowa strona danych". To był konkurs dla programistów na nowe funkcjonalności dla portalu dane.gov.pl. Powstał także film promujący portal, udostępniony na kanałach społecznościowych.

Mamy za sobą także dwa warsztaty z przedstawicielami jednostek samorządu terytorialnego, na których podzieliliśmy się doświadczeniami w otwieraniu danych. Zebraliśmy także ich dobre praktyki. W wyniku tego powstał podręcznik o otwieraniu danych. Dzisiaj udostępniliśmy go na naszej stronie dane.gov.pl. Zachęcam do zapoznania się z nim.

Odbywają się też wizyty szkoleniowo-studyjne dla pełnomocników ds. otwartości danych. Przypomnę, że w ramach programu otwierania danych publicznych jest to bardzo ważna sieć współpracy. W każdym ministerstwie, KPRM i GUS, został powołany pełnomocnik ds. otwartości danych. Jest to osoba odpowiedzialna za otwieranie danych w ministerstwie i we wszystkich jednostkach podległych i nadzorowanych. Stworzyliśmy taką właśnie sieć współpracy i pełnomocników również edukujemy.

Odbyły się właśnie dwie wizyty szkoleniowo-studyjne w ramach projektu we Francji i Hiszpanii, czyli tych państwach członkowskich, które są liderami w otwieraniu danych – zajmują pierwsze miejsca w rankingach. Dzięki nim zebraliśmy m.in. inspiracje do nowego projektu „Otwarte dane plus", o którym zaraz opowiem. Tam również dzieliliśmy się dobrymi praktykami i pozyskaliśmy wiedzę o tym, jakie rozwiązania stosowane są w innych państwach członkowskich.

Jeżeli chodzi o zadania, które jeszcze są przed nami w ramach projektu, to jest to przede wszystkim zakończenie prac nad otwarciem przez API danych z CEPIK, BESTI oraz dwóch baz NFZ. Będziemy kontynuować rozbudowę portalu dane.gov.pl. Rozbudujemy tu m.in. konto użytkownika. Będą większe możliwości śledzenia przez użytkownika różnych zbiorów, kategorii, artykułów.

Rozbudujemy także konto dysponenta danych o walidatory. Już teraz walidujemy dane, ale te walidatory będą szersze, tzn., będą sprawdzały m.in., czy plik XLS jest prawidłowo przygotowany, czy plik CSV spełnia wszystkie te wymagania, które mamy sformułowane w standardzie technicznym.

Przed nami także kontynuacja szkoleń na poziomie zaawansowanym ze standardów otwartości danych, a także realizacja działań promocyjnych. Będziemy przede wszystkim promować nasz portal dane.gov.pl, który chyba jeszcze nigdy nie był promowany w sposób profesjonalny przez agencję.

Jest to bardzo ważne, dlatego że wiemy z badań prowadzonych przez Ministerstwo Inwestycji i Rozwoju, że portal ten jest słabo rozpoznawalny wśród przedsiębiorców. Właściwie niewielu przedsiębiorców wie o istnieniu portalu. Dlatego musimy koniecznie go wypromować.

Jeżeli chodzi o efekty realizacji projektu, to już obserwujemy znaczny wzrost ilości danych udostępnianych w portalu dane.gov.pl i poprawę ich jakości. Wskaźniki przed-

stawię za moment. Poprawiła się także znacząco wiedza i poziom umiejętności dysponentów danych w zakresie otwierania danych i ich przygotowywania do udostępnienia.

Obserwujemy też w rankingach międzynarodowych, że nasza pozycja znacząco się poprawia. W rankingu OECD z ubiegłego roku awansowaliśmy na 20 miejsce, o 8 miejsc w porównaniu z rokiem 2015, natomiast w rankingu unijnym awansowaliśmy na 13 miejsce w tym roku z 22 w ubiegłym. Te ankiety międzynarodowe są bardzo szczegółowe. Dotyczą zarówno jakości danych, jak i funkcjonalności portalu, współpracy międzyinstytucjonalnej. Zatem, każda nasza odpowiedź jest wielokrotnie weryfikowana przez ekspertów. Te rankingi są naprawdę wiarygodne.

Na koniec, statystyki portalu. Od 2016 r., kiedy rozpoczęliśmy realizację programu otwierania danych, do września 2018 r. można zaobserwować znaczący wzrost liczby podmiotów udostępniających dane. Już mamy, zdaje się, ponad 120 instytucji. Liczba zasobów danych wzrosła z 3700 do 9500.

Bardzo poprawia się też jakość danych. Dane ustrukturyzowane, jeżeli chodzi o ich odsetek wśród danych, które są udostępniane w portalu, wzrósł z 73% do 89%.

Tak samo poprawia się poziom otwartości danych. Danych, na co najmniej trzecim poziomie otwartości mamy już ok. 5500.

Rośnie także liczba zasobów udostępnianych przez API. Mamy wzrost z 55 do 207 we wrześniu tego roku.

Przewidzieliśmy także kontynuację naszego projektu. Tak jak mówiłam, zebraliśmy inspiracje na naszych wizytach studyjnych. Projekt „Otwarte dane plus" został już złożony do CPPC. Czekamy na jego ocenę. Jego realizacja rozpocznie się w styczniu przyszłego roku, a potrwa 3 lata. Jego celem, podobnie jak poprzednika, jest zwiększenie ilości i poprawa jakości otwartych danych publicznych i wzrost ich ponownego wykorzystywania.

W tym projekcie otwieramy kolejne bazy przez API. Będą to: TREZOR z Ministerstwa Finansów, rejestry publiczne w turystyce z Ministerstwa Sportu i Turystyki oraz dziedzinowe bazy danych z GUS. Planujemy także rozbudowę portalu dane.gov.pl o narzędzia, które usprawnią pracę zarówno pełnomocników, jak i możliwości zarządzania i koordynowania procesem otwierania danych w urzędach.

Powstanie Laboratorium Otwartych Danych, które będzie eksperckim ośrodkiem świadczącym kompleksowe wsparcie dla urzędów, dotyczące zarówno kwestii prawnych, technicznych, jak i bezpieczeństwa. Planujemy także Akademię Otartych Danych. To będzie regularny kurs dla urzędników, dotyczący zarówno wszystkich tych aspektów otwierania danych, z których już teraz edukujemy, jak i wykorzystywania tych danych w pracy urzędniczej. Pamiętajmy bowiem o tym, że beneficjentem otwartych danych jest również administracja publiczna i ona także powinna korzystać z tych danych. Są również przewidziane działania promocyjne, które, tak naprawdę, są ważnym elementem wszystkiego tego, co robimy i pozwalają nam na wypromowanie przede wszystkim korzyści z otwierania danych wśród ich beneficjentów. To wszystko. Dziękuję.

Sekretarz stanu w MC Karol Okoński:

Na koniec, jeszcze jeden komentarz. Pewnie państwo zwrócili na to uwagę, że nasze przedsięwzięcie nie jest wyłącznie przedsięwzięciem technologicznym. Ono jak najbardziej ma aspekty edukacyjne, a efekty mogą być później zwielokrotnione przez fakt, że te dane mogą być po prostu wykorzystywane przez obywateli czy organizacje pozarządowe, przedsiębiorców, ale też samą administrację. Czyli, jest to samonapędzająca się machina, której – z czego cieszymy się – możemy nadać impet. Dziękuję.

Przewodniczący poseł Paweł Pudłowski (N):

Dziękuję bardzo pani naczelnik. Dziękuję bardzo panie ministrze. Otwieram dyskusję. Kto z państwa chciałby zabrać głos? Bardzo proszę, pani przewodnicząca Bubula.

Poseł Barbara Bubula (PiS):

Ja, słuchając państwa, równocześnie weszłam na ten portal dane.gov.pl i nic mi się tu nie otwiera. Dlaczego? Czy to nie obsługuje oprogramowania w naszych tabletach?

Sekretarz stanu w MC Karol Okoński:

Zaraz zapewnimy pani przewodniczącej indywidualną opiekę. Zaraz to sprawdzimy. Prac serwisowych na dzisiaj nie planowaliśmy, więc powinno wszystko działać.

Poseł Barbara Bubula (PiS):

Powiedzmy, że ja jestem średnim użytkownikiem Internetu. Chciałabym z ciekawości zobaczyć, które podmioty dostały zezwolenia z Ministerstwa Rolnictwa i Rozwoju Wsi na sprzedaż środków ochrony roślin. Klikam na różne linki, które są podpięte pod taką właśnie kategorię i nie otwiera się to.

Sekretarz stanu w MC Karol Okoński:

Dobrze. Za chwilę spróbujemy zrobić nawet prezentację life na tym komputerze, tylko musimy podłączyć Internet.

Przewodniczący poseł Paweł Pudłowski (N):

Poczekamy. Czy są inne uwagi lub pytania? Państwo szukają, a ja mam takie pytanie. Widzieliśmy, co się dzieje po stronie podaży danych. A co jest po stronie popytu? Jakie jest zainteresowanie użytkowników danych – przedsiębiorców czy ośrodków naukowych – korzystaniem z tych wystawionych przez państwa API i zbiorów danych? Czyli, ilu tu jest użytkowników. Czy w ogóle kontrolujemy, mierzymy, ile tych danych jest pobieranych? Czy są uwagi od nich odnośnie do jakości danych, czy braku jakichś danych, które są w przestrzeni publicznej, ale jednak nie są udostępniane? Czy były wyliczenia – może przybliżenia wyliczeń z Europy i świata – odnośnie do potencjału finansowego dla gospodarki? Wiem, że trochę na ten temat było w dokumencie przygotowanym przez Ministerstwo Przedsiębiorczości i Technologii w obszarze sztucznej inteligencji, jeśli chodzi o wykorzystanie bazy danych, ale czy państwo w tej swojej strategii, w tym swoim materiale, też analizowali, co to może oznaczać dla gospodarki jako takiej, dla PKB. Bardzo proszę, pan przewodniczący Arndt.

Poseł Paweł Arndt (PO):

Ja też miałbym kilka pytań, może dość szczegółowych. Projekt „Otwarte dane" nie jest zbyt kosztowny. On kosztuje raptem 23,5 mln zł, ale jest realizowany już mniej więcej od połowy 2017 r. i jeszcze ma to potrwać. Czy państwo byliby w stanie powiedzieć, jaka część tych środków została wykorzystana, a jaka jest jeszcze do wykorzystania?

Z tych zadań, które są zaplanowane do realizacji, najczęściej mówiliśmy tutaj, na posiedzeniach naszej Komisji, o CEPiK, dlatego nie chciałbym już dzisiaj o tym mówić, ale zainteresowało mnie inne zadanie. Co prawda, jest to zadanie realizowane przez NFZ, ale jest dość ciekawe – Ogólnopolski Informator o Czasie Oczekiwania na Świadczenia Medyczne. Proszę powiedzieć, jaki jest cel takiego informatora. Ja nie wiem, czy my chcemy frustrować naszych pacjentów i denerwować nieporadnością służby zdrowia, czy może mamy nadzieję, że placówki służby zdrowia dzięki temu informatorowi będą sprawniej działały.

Mam jeszcze jedno pytanie dotyczące szkoleń. Pani dyrektor mówiła, że bodajże 119 pracowników zostało przeszkolonych. Jak sądzę, są to pracownicy administracji centralnej. Czy takie szkolenia są przewidywane – a może są – również dla administracji samorządowej?

Przewodniczący poseł Paweł Pudłowski (N):

Bardzo dziękuję, panie przewodniczący. Czy ktoś jeszcze chciałby…? Może ktoś z gości jakiś komentarz, bo trwają nerwowe przygotowania do demonstracji online dostępu do informacji publicznej? Jeśli nie, to zamykamy listę pytań. Bardzo proszę pana ministra lub panią naczelnik o odpowiedzi.

Sekretarz stanu w MC Karol Okoński:

Dobrze. Dwie rzeczy na początek sprostuję. Pani dyrektor, strategia sztucznej inteligencji powstała w Ministerstwie Cyfryzacji, a nie w Ministerstwie Przedsiębiorczości i Technologii.

Natomiast, proszę państwa, odnośnie do tych danych, jeśli chodzi o potencjał i stopień wykorzystania, zaraz oddam głos pani dyrektor. Na pewno my to śledzimy i obserwu-

jemy, i część z tych zbiorów, które teraz są sugerowane w pierwszym i kolejnym etapie, bazują m.in. na tych sugestiach, które zbieraliśmy. Pani dyrektor zaraz to uzupełni, ja tylko odpowiem na pytanie pana przewodniczącego Arndta.

Jeśli chodzi o wydatki już poniesione w projekcie, to w tym monecie jest to 5 mln zł, czyli ¼. To wynika trochę z tego, że pierwsza faza była w dużej mierze analityczna. W niektórych przypadkach przetargi dopiero się zakończyły albo są jeszcze w trakcie, więc największy poziom wydatków będzie pod koniec projektu. Jednak w tym momencie idziemy zgodnie z harmonogramem. I zgadzam się, że z perspektywy wartości, to faktycznie ten koszt jest stosunkowo niski w porównaniu z innymi projektami.

Jeśli chodzi o NFZ i dane o kolejkach, to w momencie, kiedy była decyzja o tym, że to będzie jedna z baz, która będzie otwierana, i rozmawialiśmy z NFZ czy Ministerstwem Zdrowia, to oni to potraktowali raczej jako szansę, a nie straszenie. Może być tak, że w niektórych rodzajach zabiegów czy operacji mogą występować bardzo duże różnice pomiędzy regionami. Do tej pory istniała wyszukiwarka – do tych danych już teraz można się dostać, tylko sposób wyszukiwania jest utrudniony. Natomiast możliwość stworzenia aplikacji, która by korzystała z tych danych, pozwoli przygotować rozwiązania, które będą w stanie wskazać obywatelowi to miejsce, gdzie akurat w danym momencie ta kolejka jest krótsza. Chociaż nie ukrywam, że mogą być takie sytuacje, że generalnie w żadnym z regionów ten czas nie będzie zgodny z oczekiwaniami. Jednak NFZ i MZ na innym odcinku nad tym pracują.

Odnośnie do szkoleń, one faktycznie były na razie adresowane do administracji centralnej. Natomiast, odnośnie do tego, czy planujemy je też dla administracji samorządowej, również proszę o uzupełnienie panią dyrektor.

Dyrektor departamentu w MC Agata Miazga:

To ja płynnie przejdę do szkoleń. Program Operacyjny Polska Cyfrowa nie pozwala na to, żeby beneficjentem szkoleń były samorządy. Dlatego przewidzieliśmy dla samorządów głównie materiały edukacyjne oraz podręcznik, o którym już wspomniałam. Współpracujemy także na „miękko" z samorządami – wymieniamy się doświadczeniami, zaprosiliśmy samorządy do naszej grupy roboczej ds. otwartych danych, natomiast nie możemy bezpośrednio szkolić pracowników samorządów, dlatego że w regionalnych programach operacyjnych nie zostały przewidziane środki na te cele. POPC na to nie pozawala. Chcieliśmy to zrobić, ale, niestety, nie mogliśmy.

Jeżeli chodzi o popyt na dane, to jest to kwestia złożona przede wszystkim dlatego, że sam portal – jak już mówiłam – nie był nigdy mocno promowany ani wśród przedsiębiorców, ani wśród innych beneficjentów. Dlatego jego poziom oglądalności to ok. dwudziestu kilku tysięcy miesięcznie. Nie jest to dużo. Monitorujemy liczbę pobrań dla każdego z zasobów danych, więc wiemy, które są najbardziej popularne. Można też to sprawdzić, bo jest to widoczne. Każdy z użytkowników może sprawdzić, jaki jest poziom wykorzystania danego zasobu.

Mamy też możliwość zgłaszania na stronie zapotrzebowań na dane. Tych maili czy też informacji na razie dużo do nas nie spływa, ale myślę, że w momencie, gdy wypromujemy portal bardziej, będzie ich więcej.

Niekiedy też dostajemy zapytania obywateli, niekoniecznie w związku z portalem, tylko po prostu do naszej instytucji. Można również zgłaszać uwagi odnośnie do jakości danych. Jest taka możliwość na portalu.

Można zgłosić uwagi do tego, że dane są źle ustrukturyzowane, mają błędy. Takie informacje trafiają zarówno do nas jako administratora portalu, jak i do samego dysponenta danych, który zamieścił je w portalu.

Wartość rynku. Wartość rynku w Polsce szacuje się na 1,2 mld euro – właściwie są to szacunki, że w 2020 r. Nie mamy, tak naprawdę, dobrych badań w tym zakresie. Są pewne przymiarki, ale wszystko to zależy od tego, na ile polskie przedsiębiorstwa będą wykorzystywać dane w swojej codziennej pracy i to nie tylko do tego, żeby budować na nich aplikacje czy usługi, ale też do podejmowania decyzji biznesowych. Takich przedsiębiorstw w Polsce nie jest dużo. To jest na pewno potężne wyzwanie.

Sekretarz stanu w MC Karol Okoński:
Czy chcą państwo jeszcze krótką prezentację? Tak.

Dyrektor departamentu w MC Agata Miazga:
Wchodzimy do kategorii. Tu pojawiają się zbiory danych. Załóżmy listę…

Sekretarz stanu w MC Karol Okoński:
O jakiej kategorii danych pani przewodnicząca mówiła? Spróbujmy odtworzyć ten proces jeszcze raz.

Poseł Barbara Bubula (PiS):
Ministerstwo Rolnictwa i Rozwoju Wsi…

Dyrektor departamentu w MC Agata Miazga:
Wchodzimy na „instytucje". Wyszukujemy „Ministerstwo Rolnictwa i Rozwoju Wsi"…

Poseł Barbara Bubula (PiS):
Ja z tej głównej strony wchodziłam na te dane…

Dyrektor departamentu w MC Agata Miazga:
To wejdźmy na „Ministerstwo Rolnictwa i Rozwoju Wsi". Widzę, że Internet wolno działa. Dobrze, to jeszcze raz. Tutaj dane można pobrać – „pobierz zasób" i „otwórz".

Poseł Barbara Bubula (PiS):
A tam, gdzie jest API? Tam, na pierwszej stronie, jest, żeby w czasie rzeczywistym przy zmianach to pobrać. Czy to działa, czy nie działa?

Dyrektor departamentu w MC Agata Miazga:
W przypadku danych statycznych, takich jak ten plik XLS, pobieramy po prostu pliki XLS. Przez API możemy tylko udostępnić informacje o pliku, czyli metadane. Natomiast, sam plik pobieramy po prostu tak jak XLS, jak normalnie ze strony.

Poseł Barbara Bubula (PiS):
Czyli, po kliknięciu na to jest tylko informacja o pliku.

Dyrektor departamentu w MC Agata Miazga:
Tak. W tym przypadku, tak. W przypadku, gdy jest to zasób udostępniony przez API, mamy API.

Poseł Barbara Bubula (PiS):
Sądziłam, że taki komunikat pokaże nam dane, a nie informację o pliku, bo to właśnie…

Sekretarz stanu w MC Karol Okoński:
Przyjmujemy to jako komentarz pod kątem przyszłego rozwoju. Tak więc, jak najbardziej dziękujemy za uwagę.

Poseł Barbara Bubula (PiS):
Jest to troszkę dezinformujące.

Sekretarz stanu w MC Karol Okoński:
Dobrze. Zastanowimy się, czy można to zrobić bardziej przejrzyście. Dziękujemy.

Przewodniczący poseł Paweł Pudłowski (N):
Bardzo dziękuję. Zamykam dyskusję. Stwierdzam, że porządek dzienny posiedzenia został wyczerpany. Zamykam posiedzenie Komisji.

Protokół posiedzenia z załączonym pełnym zapisem jego przebiegu jest do wglądu w sekretariacie Komisji w Kancelarii Sejmu. Dziękuję bardzo.